Sports Medicine (2024) 54:429–446
https://doi.org/10.1007/s40279-023-01924-y

SYSTEMATIC REVIEW

The Additional Effect of Training Above the Maximal Metabolic Steady State on *V*O2peak, *W*peak and Time-Trial Performance in Endurance-Trained Athletes: A Systematic Review, Meta-analysis, and Reality Check

Michael A. Rosenblat[1] · Jem Arnold[2] · Hannah Nelson[3] · Jennifer Watt[4,5,6] · Stephen Seiler[7]

Accepted: 21 August 2023 / Published online: 22 September 2023

Abstract

Background To improve sport performance, athletes use training regimens that include exercise below and above the maximal metabolic steady state (MMSS).

Objective The objective of this review was to determine the additional effect of training above MMSS on VO_{2peak}, W_{peak} and time-trial (TT) performance in endurance-trained athletes.

Methods Studies were included in the review if they (i) were published in academic journals, (ii) were in English, (iii) were prospective, (iv) included trained participants, (v) had an intervention group that contained training above and below MMSS, (vi) had a comparator group that only performed training below MMSS, and (vii) reported results for VO_{2peak}, W_{peak}, or TT performance. Medline and SPORTDiscus were searched from inception until February 23, 2023.

Results Fourteen studies that ranged from 2 to 12 weeks were included in the review. There were 171 recreational and 128 competitive endurance athletes. The mean age and VO_{2peak} of participants ranged from 15 to 43 years and 38 to 68 mL·kg^{-1}·min^{-1}, respectively. The inclusion of training above MMSS led to a 2.5 mL·kg^{-1}·min^{-1} (95% CI 1.4–3.6; $p < 0.01$; $I^2 = 0\%$) greater improvement in VO_{2peak}. A minimum of 81 participants per group would be required to obtain sufficient power to determine a significant effect (SMD 0.44) for VO_{2peak}. No intensity-specific effect was observed for W_{peak} or TT performance, in part due to a smaller sample size.

Conclusion A single training meso-cycle that includes training above MMSS can improve VO_{2peak} in endurance-trained athletes more than training only below MMSS. However, we do not have sufficient evidence to conclude that concurrent adaptation occurs for W_{peak} or TT performance.

1 Introduction

To improve sport performance, athletes use training regimens that include exercise below and above the maximal metabolic steady state (MMSS) [1–3]. MMSS is considered to be the exercise intensity that coincides with the maximal sustainable oxidative metabolic rate [4]. MMSS is a theoretical threshold that separates the heavy and severe intensity domains (Fig. 1), where exercise performed within the heavy domain allows for the attainment of a quasi-steady state systemic oxygen consumption ($\dot{V}O_2$) [5]. A metabolic steady state is not attainable in the severe domain [5]. Since a steady state cannot be achieved at intensities above MMSS, exercise is typically performed as intervals of work and rest periods to increase the total time spent above MMSS within a single session.

A systematic review and meta-analysis by Rosenblat et al. [6] was conducted to determine the optimal method to program training above MMSS to maximize gains in time-trial (TT) performance. The results of the review showed that exercise performed at any intensity above MMSS would lead to similar improvements in TT performance in endurance-trained athletes. Increasing the total duration of training above MMSS was found to lead to better performance outcomes [6]. However, trying to maximize the time spent above MMSS in every training session could lead to delayed fatigue, acute performance impairment, and increased risk of illness [7–9]. Therefore, athletes should focus on optimizing, not maximizing the time spent above MMSS both within a single exercise session and across a training cycle.

Extended author information available on the last page of the article

Key Points

A single meso-cycle that includes endurance training above the maximal metabolic steady state (MMSS) will improve $\dot{V}O_{2peak}$ in endurance-trained athletes to a greater extent than training only below MMSS.

Sources of clinical heterogeneity such as differences in participant characteristics, exercise mode, and training programming variables do not influence the adaptive responses that occur with adding exercise above MMSS to an endurance training program in endurance-trained athletes.

All the studies included in the review were underpowered to detect a significant effect across outcomes because of small sample sizes, which remains a common and critical limitation of laboratory-based training intervention studies.

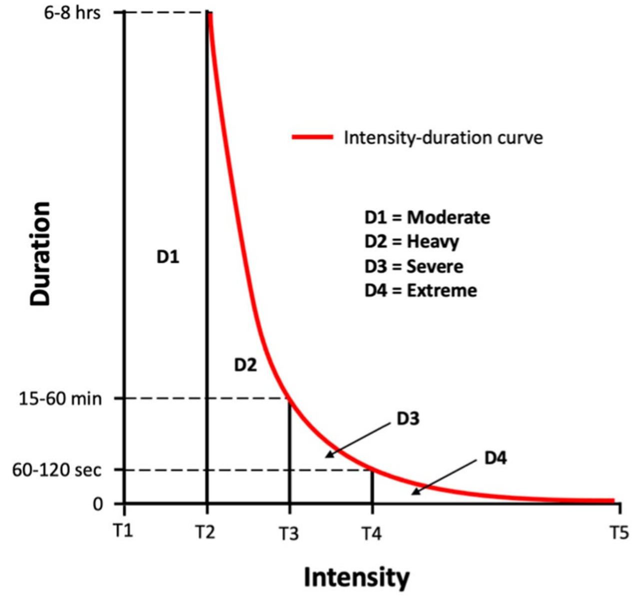


Fig. 1 Four domain model. Threshold 1 (T1) marks the onset of exercise. Threshold 2 (T2) represents the gas exchange threshold (GET), the first ventilatory threshold (VT_1), or the first lactate threshold (LT_1). Threshold 3 (T3) represents the maximal metabolic steady state (MMSS) which can be determined by the respiratory compensation point (RCP), second ventilatory threshold (VT_2), the second lactate threshold (LT_2), maximal lactate steady state (MLSS), or critical power (CP). Threshold 4 (T4) indicates the maximal aerobic power (MAP), which is the highest power/speed that allows for the attainment of maximal oxygen consumption ($\dot{V}O_{2max}$). Threshold 5 (T5) indicates the peak power output achieved from a Wingate test

For endurance athletes performing high volumes of training, not every session can include high-intensity exercise. Findings from prospective observational studies indicate that endurance athletes follow a training program that consists of a relatively low percentage of high-intensity training sessions [1–3]. To maximize improvements in TT performance, it may be most beneficial for athletes to perform the majority (~80%) of their training at relatively low intensities (i.e., below the first lactate or first ventilatory threshold) with only 10–20% at intensity above MMSS [10].

TT tests may be the most appropriate, controllable method to measure performance in endurance athletes since TTs reasonably simulate the demands of racing. The most common type of TT test requires an individual to complete a set distance (e.g., 40 km) in the shortest time possible. The integrated performance model proposed by Joyner and Coyle [11] suggests that race and/or TT performance would be positively impacted by increases in maximal oxygen consumption ($\dot{V}O_{2max}$), fractional utilization of $\dot{V}O_{2max}$ at MMSS, gross mechanical efficiency, and to a lesser extent depending on duration, anaerobic capacity.

$\dot{V}O_{2max}$ is a hallmark indicator of aerobic capacity, measured as the maximal rate at which oxygen is utilized by an individual [12]. It is most frequently obtained by measuring respiratory gas exchange during a ramp or incremental exercise test (IET) to maximal task tolerance. The strongest confirmational marker of attainment of $\dot{V}O_{2max}$ is a plateau in oxygen consumption ($\dot{V}O_2$ rises ≤ 150 mL·min^{-1} despite a continued increase in work rate) [13]. When a plateau is not observed, the test results would be defined as peak oxygen consumption ($\dot{V}O_{2peak}$) [14]. In addition to $\dot{V}O_{2peak}$, the highest power (W_{peak}) or velocity (V_{peak}) is also recorded at the end of an IET and represents an integration of aerobic capacity, anaerobic capacity, and gross mechanical efficiency [15]. Moving forward in this review, the terms $\dot{V}O_{2max}$ and $\dot{V}O_{2peak}$ will be referred to exclusively as $\dot{V}O_{2peak}$, and W_{peak} and V_{peak} exclusively as W_{peak}, unless otherwise specified. Both $\dot{V}O_{2peak}$ [16] and W_{peak} [17] have been shown to improve with training above MMSS.

$\dot{V}O_{2peak}$ is influenced by several central (e.g., plasma volume, left ventricular mass, stroke volume, etc.) and peripheral (e.g., capillary density, mitochondria, etc.) physiological factors [12]. Changes in mitochondrial content and respiratory function have been shown to be intensity dependent, with greater improvements following training above MMSS compared with below MMSS [18, 19]. Previous findings suggest that there is no significant difference between interventions that incorporate training above versus below MMSS on other factors including plasma volume [20–23], left ventricular mass [23–25], stroke volume [21, 23–25], cardiac output [21, 23–25], or capillary density [26]. It is important to note that these studies included relatively small sample sizes (12–29 participants), which may limit the ability to detect a significant effect. In addition, the studies were primarily conducted in untrained individuals, likely to aid in participant recruitment. Therefore, the observed improvements that occurred with both interventions may simply be a result of the addition of a structured endurance training program providing sufficient stimulus for adaptations to occur.

There are systematic reviews with meta-analyses that compare training above versus below MMSS on $\dot{V}O_{2peak}$ in untrained [27–29] and clinical populations [30]. These studies showed that training exclusively above MMSS produced greater improvements in $\dot{V}O_{2peak}$ when compared with exclusively below MMSS. Since training status (i.e., untrained versus trained) has been shown to influence changes in $\dot{V}O_{2peak}$ [16], the results cannot be generalized to endurance-trained athletes. Furthermore, as discussed, it is common for athletes to include a large volume of low-intensity training within their training programs. Therefore, a comparison of exclusively training above versus exclusively below MMSS in endurance athletes would have low ecological validity.

It is currently unclear how the addition of exercise performed above MMSS can influence performance when added to a training program that already consists of low-intensity training (< MMSS). A systematic review with meta-analysis will help to establish if there are additional benefits to adding exercise above MMSS to an existing low-intensity training program and the magnitude of the improvement. Currently, there are no systematic reviews that examine the additional effect of training above MMSS on endurance performance in endurance-trained athletes. Therefore, the objective of this systematic review and meta-analysis was to determine the additional effect of endurance training above MMSS on $\dot{V}O_{2peak}$, W_{peak}, and TT performance in endurance-trained athletes.

2 Methods

2.1 Eligibility Criteria

Studies were included in the systematic review if they (i) were published in peer reviewed, academic journals; (ii) were published in English; (iii) were prospective; (iv) included trained participants; (v) included an intervention group that contained training above and below MMSS; (vi) included a comparator group that only performed training below MMSS; and (vii) reported results for $\dot{V}O_{2peak}$ relative to body mass, W_{peak}, or TT performance (time to complete a set distance). Participants were considered trained if they participated in a structured training program that was specific to a mode of exercise, as previously described [6].

Studies were excluded for the following reasons: (i) the full-text article was unavailable; (ii) participant training status was not described; (iii) the study included individuals from multiple sports; (iv) the intervention was not clearly described; (v) the study included nutritional interventions (supplements, hydration, fed state, etc.); (vi) participants were subject to changes in environmental conditions (heat, cold, altitude, hypoxia, hyperoxia, etc.); (vii) the study included potential ergogenic devices/modalities (cooling vests, compression garments, etc.); (viii) the study included pharmacological agents; (ix) the study included multiple meso-cycles or changes in intensity distribution; (x) training exercise mode was different from the method used to determine $\dot{V}O_{2peak}$, W_{peak}, or TT performance; or (xi) the study did not include group means with accompanying measurements of uncertainty for $\dot{V}O_{2peak}$ relative to body mass, W_{peak}, or TT performance as time to complete a set distance, at baseline and following the intervention.

2.2 Information Sources

Two databases, Medline (Ovid) and SPORTDiscus (EBSCOhost), were electronically searched across all publication years (i.e., from inception) up to and including February 23, 2023. Reference lists from relevant systematic reviews identified while screening the titles and abstracts were also searched for additional eligible articles.

2.3 Search Strategy

The search strategy followed PICOS guidelines to ensure that a broad and inclusive search was conducted. The search strategy included a list of synonyms and tenses for all relevant participant characteristics, interventions, and outcomes. The search limits were set to the following: titles, abstracts, English language, and academic journal articles. A line-by-line search strategy is available in the electronic supplementary material (ESM), Appendix S1.

2.4 Selection Process

Reviewer pairs independently completed two levels of article screening (1. title and abstract and 2. full-text) and abstracted data from full-text articles. Non-resolvable disagreements were adjudicated by a third person.

2.5 Data Collection Process

One reviewer completed a charting exercise prior to data abstraction to identify interventions and outcomes reported in the studies and to inform the data abstraction plan. The reviewer used established imputation methods for missing measures of uncertainty and data transformation. When data were only available in graphical form (i.e., a bar or line graph), we used PlotDigitizer [31] to extract the data points. Where missing data could not be determined, the authors of the respective studies were contacted via email to request the individual and/or aggregate data. Two reviewers independently extracted the data from each of the studies. The two reviewers compared their findings to check for errors. Any inconsistencies were discussed and addressed by the two reviewers. A third reviewer was consulted if the first two reviewers were unable to reach agreement.

When multiple studies reported data from the same study population, the publication with the most complete outcome data was considered the primary publication; otherwise, the publication with the largest sample size was considered the primary publication. If a companion publication reported data for an additional outcome of interest, then the data were extracted, but publication details (e.g., sample size) were abstracted from the primary publication only.

2.6 Data Items

Participant baseline data (means, SD, and/or SEM) were extracted for the following participant characteristics: age, sex, training status, body mass, height, body mass index (BMI), and $\dot{V}O_{2peak}$. Training status was classified as competitive if participants competed at a high-performance level (i.e., varsity, provincial/state, national, international, or professional); anything else was classified as recreational.

Extracted intervention characteristics included training mode (i.e., cycling, running, etc.), exercise type (continuous versus interval), weeks, frequency, sets, repetitions, work-bout intensity and duration, intensity domain, recovery mode and duration, and load progression variables. Progressive overload was determined if an intervention included an increase in any one of the programming variables over the course of the intervention (e.g., an increase in intensity).

The correction factor, $1.8596 \cdot \text{test duration (min)}^{-0.242}$, based on the model proposed and validated by Morton [32], and further validated by Adami et al. [33], was used to standardize W_{peak} obtained from incremental testing protocols that exceeded 12 min in duration. Exercise intensity was then converted to a percentage of the adjusted W_{peak} for all instances where it was expressed as absolute power (watts) or speed ($km{\cdot}h^{-1}$).

Training intensity domains were determined using the Four Domain Model, which divides exercise into the moderate, heavy, severe, and extreme domains as described by Poole and Jones [5] and shown in Fig. 1. The first lactate turn-point/threshold (LT_1), the gas exchange threshold (GET), and the first ventilatory threshold (VT_1) were all used to define the transition between the moderate and heavy domains; all three of these have been shown to occur at the same percentage of $\dot{V}O_{2peak}$ [34, 35]. The second ventilatory threshold (VT_2), respiratory compensation point (RCP), critical power (CP), second lactate turn-point/threshold (LT_2), and maximal lactate steady state (MLSS) were all used as methods to estimate MMSS and define the transition between the heavy and severe domains. The adjusted value for W_{peak} was used to separate the severe and extreme domains.

Each estimate of MMSS is defined by different physiological criteria, and as such there are expected small differences in the intensity and workload at which MMSS will be calculated. For instance, MLSS has been shown to occur at a lower $\dot{V}O_2$ [36] and a lower power/speed [37] than CP. Furthermore, exercise time to exhaustion varies across the techniques, with sustainable durations lasting for as little as 15 min at CP [38–40] or as long as 60 min at MLSS [41–44]. Since these methods are all commonly used to determine MMSS, they were used interchangeably to classify exercise in this review. However, we acknowledge that each method will produce a different value for MMSS, and that MMSS itself likely represents a transition event rather than a discrete threshold [45].

To account for any differences in the measurement techniques used to program intensity among the studies (e.g., heart rate [HR], power/speed, $\dot{V}O_{2peak}$), the findings from Iannetta et al. [46] were used as a guide to place the interventions in the correct intensity domain. Specifically, training performed at or below 75% of maximal heart rate (HR_{max}), 40% W_{peak}, or 60% $\dot{V}O_{2peak}$ was allocated to the moderate domain. Training above those intensities, but at or below 85% HR_{max}, 60% W_{peak}, or 80% $\dot{V}O_{2peak}$, was allocated to the heavy domain.

The mean, SD, and/or SEM for outcomes including $\dot{V}O_{2peak}$, W_{peak}, and TT performance were extracted from each study. If studies included a preparation/wash-in training period, then data were collected from three timepoints: baseline (i.e., prior to the preparation training period), pre-intervention (i.e., after the preparation training but prior to the intervention), and post-intervention (i.e., following the training period). Data for $\dot{V}O_{2peak}$ were extracted if the results were expressed relative to body mass ($mL{\cdot}kg^{-1}{\cdot}min^{-1}$). When relative measurements were not provided for $\dot{V}O_{2peak}$, data were extracted if absolute measurements of $\dot{V}O_{2peak}$ (expressed in $mL{\cdot}min^{-1}$) and body mass (kg) were available for all assessment timepoints performed in the respective studies. Data for TT were extracted for results expressed as time to complete a set distance.

2.7 Study Risk of Bias Assessment

The Cochrane Collaboration Risk of Bias 2.0 Tool was used to determine the level of bias in all included intervention studies [47]. Two reviewers independently assessed the individual articles with a third reviewer to resolve discrepancies.

2.8 Effect Measures

Physiological (i.e., $\dot{V}O_{2peak}$) and performance (i.e., W_{peak} and TT performance) outcome measurements were evaluated by using the standardized mean difference (SMD) between intervention groups at follow up. The SMD indicates the difference in the mean of each group relative to the group SD for the outcome measure, as a measure of how many SDs the intervention group differed compared with the control group at follow up [48]. It was used in place of expressing outcomes in their original units to synthesize the results for W_{peak}, which was expressed in different units (e.g., watts versus speed), and TT performance, which included different distance events (e.g., 2 km vs 40 km). Hedges' g was used to account for small sample size bias [49]. In addition to SMD, $\dot{V}O_{2peak}$ was also evaluated as a mean difference (MD) in its original units (mL·kg^{-1}·min^{-1}) since it is a measure that was consistent across studies.

2.9 Synthesis Methods

Data synthesis was performed for the SMD for $\dot{V}O_{2peak}$, W_{peak}, and TT performance between intervention groups at follow up. The SMD was calculated by using the sample size, means, and SDs for the respective outcomes using the Metafor package (version 3.8-1) [50]. Care was taken to prevent double counting of participants when a study contained three or more groups. Therefore, groups were combined to create a single pair-wise comparison, as described in the Cochrane Handbook [48]. The groups were combined as follows: (i) groups that only included exercise performed at intensities below MMSS were combined into a single group, and (ii) groups that included exercise bouts performed above MMSS were combined into a single group.

Data conversions were performed to determine values for missing data as well as to standardize scores to allow for a consistent interpretation of the results. In cases where relative measurements were not available for $\dot{V}O_{2peak}$, data transformation was conducted if aggregate data were available for absolute measurements of $\dot{V}O_{2peak}$ (mL·min^{-1}) and body mass (kg) to obtain values for $\dot{V}O_{2peak}$ relative to body mass. Data expressed using the SEM were converted to a SD using the following formula: $SD = SEM \bullet \sqrt{n}$; where n represents the sample size of the group [51].

All statistical analyses were performed using R (version 4.2.2) [52]. The code used to conduct the analyses is available in the ESM, Appendix S2. The MD for $\dot{V}O_{2peak}$ and the SMD for $\dot{V}O_{2peak}$, W_{peak}, and TT performance were pooled using the Metafor package (version 3.8-1) [50] by means of a random effects model and the DerSimonian–Laird estimator [53]. If a significant difference between intervention groups was present, the SMD was used to determine the statistical power of the individual studies included in the meta-analysis using the Metameta package [54]. In addition, the PWR package (version 1.3-0) based on the power calculations by Cohen [55], was used to determine the sample size that would be required for an individual study to determine the estimated SMD from the meta-analysis.

The I^2 statistic was used to describe the degree of statistical heterogeneity by determining the percentage of the total variation in the estimated effect across studies [56]. To determine the cause of statistical heterogeneity, subgroup analyses and meta-regressions were performed for five participant characteristics (age, sex, training status, BMI, and baseline $\dot{V}O_{2peak}$) and four training characteristics (exercise mode, frequency, total sessions, and total training weeks). Leave-one-out diagnostics [57] were performed to determine if there was a single study that strongly influenced the estimated effect. Sensitivity analyses were performed to determine the influence of including non-randomized studies, group stratification by participant characteristics, the effect of including comparator groups that altered their training, combining results for $\dot{V}O_{2max}$ with $\dot{V}O_{2peak}$, and the removal of outlier studies.

Tables were used to describe the study characteristics and results for $\dot{V}O_{2peak}$, W_{peak}, and TT performance from the individual studies. A flow diagram was used to describe the article screening and selection process. A figure that contained a forest plot and accompanying traffic light images was used to describe the pooled analysis of the results for $\dot{V}O_{2peak}$, W_{peak}, and TT performance and the risk of bias of the individual studies, respectively.

2.10 Reporting Bias Assessment

Egger's test was used to quantitatively assess for small study effects [58]. The presence of publication bias was assessed by comparing the pooled effect with an adjusted pooled effect estimated by the Vevea and Hedges' Weight-Function Model for Publication Bias [59] using the Weightr package (version 2.0.2). The selection model provided a likelihood ratio which was considered significant at $p < 0.10$.

3 Results

The datasets used in the quantitative analysis are available in the ESM, Appendix S3.

3.1 Study Selection

The databases Medline (Ovid) and SPORTDiscus (EBSCO-host) yielded a total of 3276 results. The search results for both databases can be found in Appendix S1 (see ESM). Following the removal of 882 duplicates, 2394 titles and abstracts were screened. A total of 63 full-text articles were identified for retrieval and screened for eligibility. Of these, 14 studies were included in the quantitative analysis (Fig. 2). The rationale for eligibility for each of the studies can be found in Appendix S1 (see ESM).

3.2 Study Characteristics

The characteristics of the individual studies are presented in Table 1. The 14 training studies ranged from 2 to 12 weeks in duration. Seven of the 14 studies included a randomization process [60–66] and 2 of 14 studies stratified participants by baseline characteristics [67, 68]. There were 171 recreational (female = 12) and 128 competitive (female = 39) cyclists, rowers, and runners. The mean age and $\dot{V}O_{2peak}$ of the participant groups ranged from 15 to 43 years and 38 to 68 mL·kg^{-1}·min^{-1}, respectively.

Five studies included three or more intervention groups [66, 67, 69–71]. The groups within those studies were combined as described in the methodology. One group (4 × 16 min) from the study by Seiler et al. [71] was excluded from the pooled analysis because it was unclear if the participants had consistently trained above or below MMSS when executing the 4 × 16 min work prescription.

Twelve of the 14 included studies measured changes in $\dot{V}O_{2peak}$ [60–62, 64, 65, 67–73], 10 of the 14 studies measured changes in W_{peak} [60, 62, 65–71, 73], and 6 of the 14 studies examined changes in TT performance [60, 63, 64, 66–68]. Only one study included a testing familiarization session where participants completed the same protocol used to determine the outcome results [69].

3.3 Risk of Bias in Studies

The full results of the risk of bias of the individual studies can be found in Fig. 3. Seven of twelve studies for $\dot{V}O_{2peak}$, three of seven studies for W_{peak}, and one of four studies for TT performance were considered to have a high risk of bias. Two studies that included a randomization process were subsequently categorized as non-randomized due to changes in group allocation post-randomization. One of the studies allowed participants to choose to be in the control group if they were not interested in taking part in the intervention [69] and the other study allowed two participants to switch groups post-randomization due to scheduling conflicts [71].

3.4 Results of Individual Studies

The results for $\dot{V}O_{2peak}$, W_{peak}, and TT performance of the individual studies are presented in Tables 2, 3, and 4, respectively. Plot digitizer was used to determine the mean effect and standard deviation for TT performance in two studies [64, 67]. There was no difference between the intervention and comparator groups at baseline for $\dot{V}O_{2peak}$ or TT performance in any of the studies. There was a significant difference in W_{peak} at baseline for one study [70], which was subsequently excluded from the analysis.

3.5 Results of Syntheses

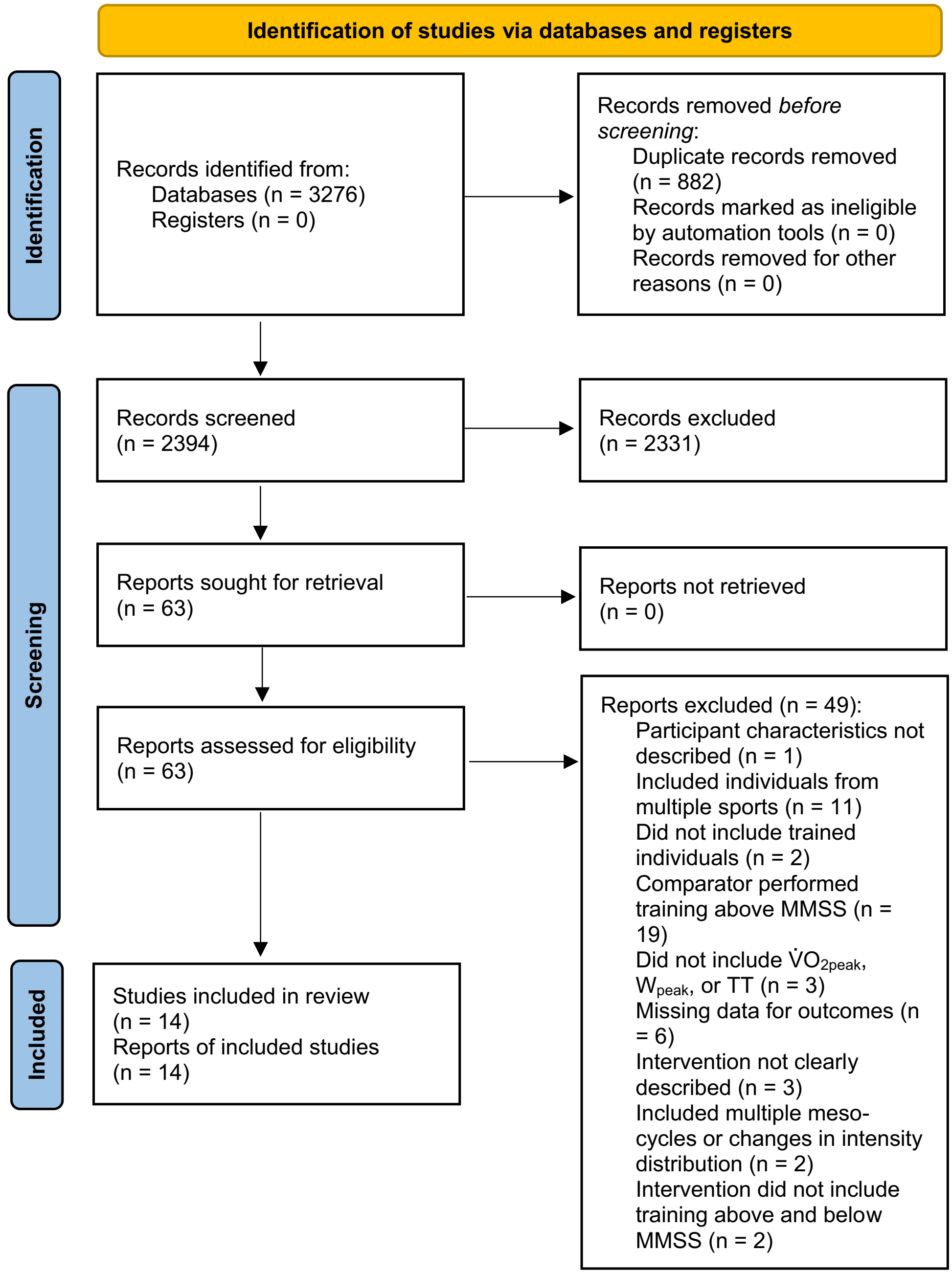
3.5.1 $\dot{V}O_{2peak}$

Twelve studies were included in the pooled estimate [60–62, 64, 65, 67–73]. There was a significant difference between intervention groups at follow up that favoured adding training above MMSS to lower intensity training for the improvement of $\dot{V}O_{2peak}$ (SMD 0.44, 95% confidence interval [CI] 0.19–0.70; $p < 0.01$). There was no evidence of statistical heterogeneity ($I^2 = 0\%$) in the pooled analysis. The MD in $\dot{V}O_{2peak}$ in original units was 2.5 mL·kg^{-1}·min^{-1} (95% CI 1.4–3.6; $p < 0.01$, $I^2 = 0\%$); see Fig. 3 for a forest plot of the synthesized results expressed as SMD. If the estimated SMD is the expected effect size, then the statistical power of the studies included in the meta-analysis would only be 16.5% (95% CI 14.4–18.5). To obtain a power of 80% to detect a SMD of 0.44 for $\dot{V}O_{2peak}$, the minimum sample size would need to be 81 participants per group in the evaluated intervention studies.

Although there was no evidence of statistical heterogeneity, stratified and meta-regression analyses were performed to evaluate the influence of clinical heterogeneity (i.e., elements of PICOS) on the overall effect. Regarding participant characteristics, there was no significant effect for modifiers, including training status ($p = 0.27$), sex ($p = 0.41$), age ($p = 0.19$), BMI ($p = 0.27$), or baseline $\dot{V}O_{2peak}$ ($p = 0.42$). There was also no significant effect for training variables including exercise mode ($p = 0.94$), exercise frequency ($p = 0.95$), total sessions ($p = 0.46$), or total weeks ($p = 0.31$).

Sensitivity analyses showed that there was no effect of including quasi-experimental studies ($p = 0.42$), stratifying participants by baseline characteristics ($p = 0.62$), including comparator groups that altered their usual training program

Fig. 2 PRISMA flow diagram. *MMSS* maximal metabolic steady state, *TT* time trial, $\dot{V}O_{2peak}$ peak oxygen consumption, W_{peak} peak power output from an incremental exercise test for cycling in watts or running velocity in km·h^{-1}



($p = 0.52$), or combining results for $\dot{V}O_{2peak}$ and $\dot{V}O_{2max}$ ($p = 0.17$).

3.5.2 W_{peak}

Seven studies that included W_{peak} were included in the pooled analysis [60, 62, 65, 66, 68, 71, 73]. There was no significant difference between intervention groups at follow up (SMD 0.12, 95% CI −0.23 to 0.48; $p = 0.49$). There was also no evidence of statistical heterogeneity ($I^2 = 0\%$) in pooled analysis; see Fig. 3 for a forest plot of the synthesized results.

There was no significant effect for modifiers including training status ($p = 0.61$), sex ($p = 0.98$), age ($p = 0.83$), BMI ($p = 0.64$), or baseline $\dot{V}O_{2peak}$ ($p = 0.86$). There was also no effect of exercise mode ($p = 0.95$), exercise frequency ($p = 0.54$), total sessions ($p = 0.92$), or total weeks ($p = 0.82$).

One study was excluded because the adjusted W_{peak} could not be calculated [69]; a second study was excluded because the groups were not similar at baseline [70]; and a third study was determined to be an outlier and was therefore excluded [67]. A sensitivity analysis was performed to determine the

Table 1 Study characteristics

Study	Study design		Participants					Intervention			Outcome
	Design	Weeks	Sex	Sport	Training status	Age Mean ± SD	$\dot{V}O_{2peak}$ Mean ± SD	Group	n	Description	
Born et al. [72]	NRCT	3	Male	Running	Recreational	25.0 ± 3.6	50.5 ± 4.6	HIIT	16	UT, D3	$\dot{V}O_{2peak}$
								LSD	12	UT, D1	
Driller et al. [60]	RCT	4	Both	Rowing	Competitive	19.0 ± 2.0	54.0 ± 6.9	HIT	10	UT, D3	$\dot{V}O_{2peak}$, W_{peak}, TT
								CT	10	UT, D2	
Esfarjani and Laursen [67]	NRCT	10	Male	Running	Recreational	19.0 ± 2.0	51.6 ± 2.9	G_1	6	D2, D3	$\dot{V}O_{2peak}$, W_{peak}, TT
								G_2	6	D2, D4	
								G_{CON}	5	D2	
Ferley et al. [69]	NRCT	6	Both	Running	Recreational	27.4 ± 3.8	60.9 ± 8.5	G_{Flat}	12	D2, D3	$\dot{V}O_{2peak}$, W_{peak}
								G_{Hill}	12	D2, D3	
								G_{CON}	8	UT	
Hebisz et al. [61]	RCT	8	Both	Cycling	Competitive	24.8 ± 4.4	59.6 ± 7.2	I	13	D1, D3, D4	$\dot{V}O_{2peak}$, W_{peak}
								Endurance	13	D1, D2	
Hebisz et al. [70]	NRCT	8	Male	Cycling	Competitive	21.1 ± 5.4	57.2 ± 5.2	E1	10	D1, D3, D4	$\dot{V}O_{2peak}$, W_{peak}
								E2	10	D1, D3, D4	
								C	7	D1, D2	
Hottenrott et al. [62]	RCT	12	Both	Running	Recreational	43.4 ± 6.9	37.9 ± 4.8	AW	14	D2, D3, D4	$\dot{V}O_{2peak}$, W_{peak}
								WE	16	D2	
Ingham et al. [63]	RCT	12	Male	Rowing	Competitive	24.0 ± 1.6	60.7 ± 4.3	Mix	9	D2, D3	TT
								Low	9	D2	
Kirchenberger et al. [64]	RCT	8	Male	Rowing	Competitive	15.3 ± 1.3	59.0 ± 3.9	IG	10	D1, D3	$\dot{V}O_{2peak}$, TT
								CG	7	D1	
Laursen et al. [73]	NRCT	2	Male	Cycling	Competitive	23.5 ± 3.5	67.5 ± 3.7	HIT	7	UT, D3	$\dot{V}O_{2peak}$, W_{peak}
								CON	7	UT	
Michalik et al. [65]	RCT	7	Male	Running	Recreational	31.9 ± 6.4	49.6 ± 5.0	Experiment	6	D2, D3	$\dot{V}O_{2peak}$, W_{peak}
								C	6	D2	
Seiler et al. [71]	NRCT	7	Both	Cycling	Recreational	40.9 ± 7.3	50.4 ± 5.8	4 × 4 min	9	D1, D3	$\dot{V}O_{2peak}$, W_{peak}
								4 × 8 min	9	D1, D3	
								4 × 16 min	9	D1, D2	
								Low	8	D1	
Silva et al. [68]	NRCT	4	Male	Running	Recreational	33.5 ± 7.5	55.6 ± 7.7	HIT	8	UT, D3	$\dot{V}O_{2peak}$, W_{peak}, TT
								CON	8	UT	
Stepto et al. [66]	RCT	3	Male	Cycling	Competitive	26.3 ± 4.6	63.2 ± 8.2	12 × 30 s	4	UT, D4	TT, W_{peak}
								12 × 60 s	3	UT, D3	
								12 × 2 min	4	UT, D3	

Table 1 (continued)

Study	Study design		Participants					Intervention			Outcome
	Design	Weeks	Sex	Sport	Training status	Age Mean ± SD	$\dot{V}O_{2peak}$ Mean ± SD	Group	*n*	Description	
								8 × 4 min	4	UT, D3	
								4 × 8 min	4	UT, D2	

AW after work, *C*, *CG*, *CON*, G_{CON} control group, *CT* traditional training, *D1* moderate domain, *D2* heavy domain, *D3* severe domain, *D4* extreme domain, *E1*, *E2* concurrent high-intensity interval training, sprint interval training, and continuous training, G_1 group 1, G_2 group 2, G_{Flat} level-grade interval training, G_{Hill} uphill interval training, *HIIT*, *HIT* high-intensity interval training, *I* interval and endurance training, *IG* intervention group, *Low* low-intensity, *LSD* long-slow distance, *Mix* mixed intensity, *NRCT* non-randomized controlled trial, *RCT* randomized controlled trial, *SD* standard deviation, *TT* time trial, *UT* usual training, $\dot{V}O_{2peak}$ peak oxygen consumption, *WE* weekend, W_{peak} peak power output from an incremental exercise test for cycling in watts or running velocity in km·h^{-1}

influence of removing the outlier study. There was a significant difference between subgroups (pooled effect vs the outlier study) (SMD 2.6, 95% CI 1.19–4.06; $p<0.001$). The SMD decreased from 0.39 (95% CI −0.17 to 0.96; $p<0.05$) to 0.12 (95% CI −0.23 to 0.48; $p=0.49$) and the statistical heterogeneity (I^2) decreased from 61 to 0% when the outlier study was removed. There was no effect of including quasi-experimental studies ($p=0.88$), including comparator groups that altered their usual training program ($p=0.88$), or stratifying participants by baseline characteristics ($p=0.28$).

3.5.3 TT Performance

Five studies were included in the pooled effect estimate [60, 63, 64, 66, 68]. There was no significant difference between intervention groups at follow up (SMD −0.10, 95% CI −0.54 to 0.33; $p=0.64$). There was also no evidence of statistical heterogeneity ($I^2=0\%$) in the pooled analysis; see Fig. 3 for a forest plot of the synthesized results.

There was no significant effect of participant characteristics including training status ($p=0.85$), sex ($p=0.65$), age ($p=0.49$), BMI ($p=0.76$), or baseline $\dot{V}O_{2peak}$ ($p=0.39$). There was also no effect of exercise mode ($p=0.92$), exercise frequency ($p=0.83$), total sessions ($p=0.42$), or total weeks ($p=0.47$). Furthermore, performance gains were not specific to time-trial duration ($p=0.67$).

There was one outlier study that was excluded from the analysis [67]. A sensitivity analysis showed that there was a significant difference between subgroups (pooled effect vs the outlier study) ($p<0.04$). There was no effect of including quasi-experimental studies ($p=0.85$), including comparator groups that altered their usual training program ($p=0.94$), or stratifying participants by baseline characteristics ($p=0.85$).

3.6 Reporting Biases

Egger's test for funnel plot asymmetry indicated that there was no evidence of small sample size bias in the results for the MD in $\dot{V}O_{2peak}$ ($z=0.32$, $p=0.75$), or the SMD in W_{peak} ($z=0.86$, $p=0.39$) or TT performance ($z=-0.38$, $p=0.71$).

The Vevea and Hedges' Weight-Function Model for Publication Bias showed that there was no evidence of publication bias in the results for $\dot{V}O_{2peak}$ ($X^2=0.36$, $p=0.55$), W_{peak} ($X^2=0.72$, $p=0.40$), or TT performance ($X^2=0.16$, $p=0.69$).

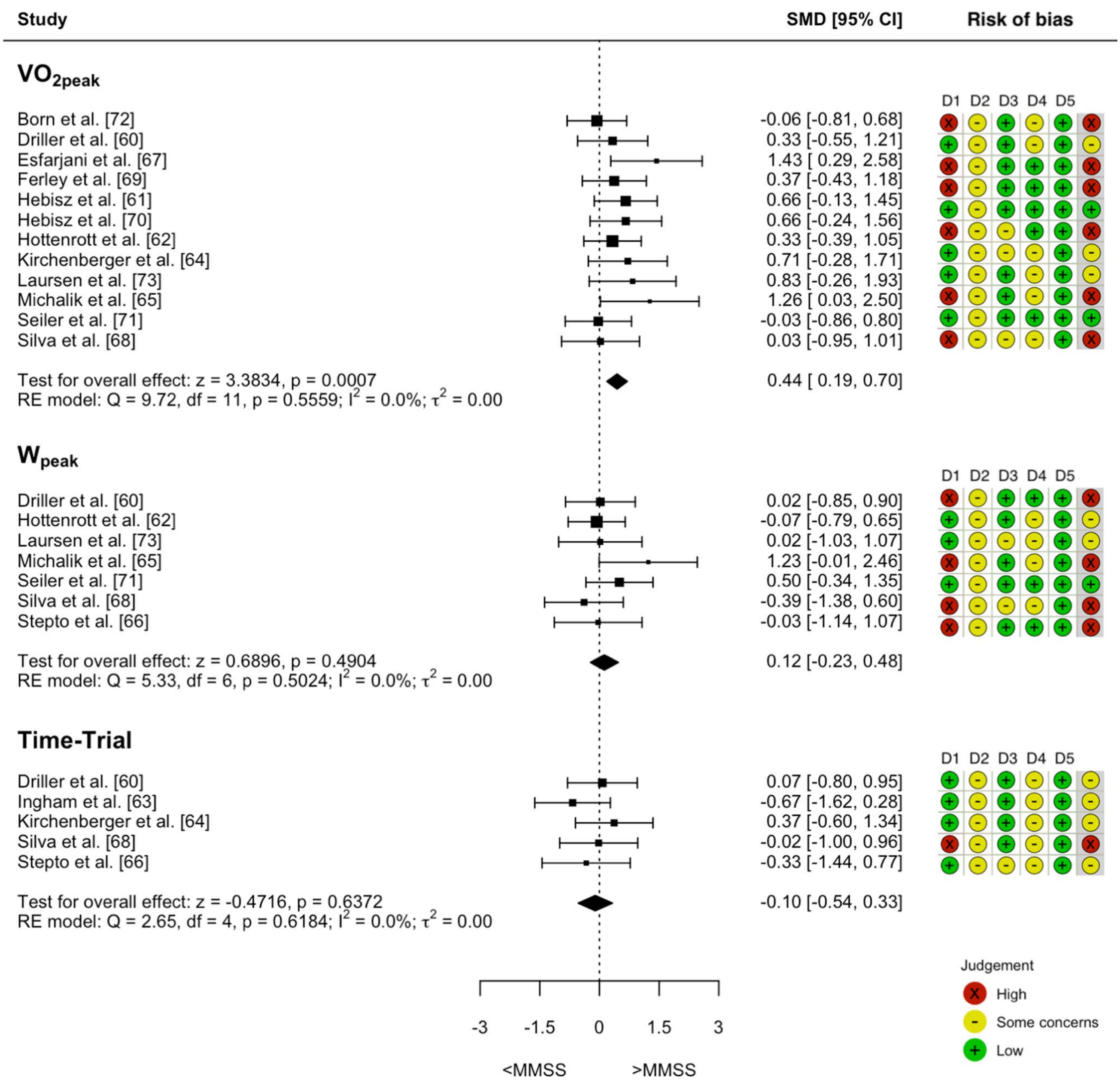


Fig. 3 Forest plot of the standardized mean difference (SMD) in maximal oxygen consumption ($\dot{V}O_{2peak}$), peak power output from an incremental exercise test (W_{peak}), and time-trial performance between training below maximal metabolic steady state (MMSS) and training above MMSS at follow up. A positive value favours training above MMSS and a negative value favours training below MMSS. Bias due to randomization (D1), bias due to deviations from intended intervention (D2), bias due to missing data (D3), bias due to outcome measure (D4), bias due to selection of reported result (D5). The traffic lights located in the grey bars depict the overall risk of bias score for each study

4 Discussion

4.1 General Interpretation of the Results

This is the first systematic review and meta-analysis that examined the additional effect of endurance training above MMSS on $\dot{V}O_{2peak}$, W_{peak}, or TT performance in endurance-trained athletes. The available data showed that adding training above MMSS led to greater improvements in $\dot{V}O_{2peak}$ (~ 2.5 mL·kg^{-1}·min^{-1}). Importantly, we found

Table 2 Results of individual studies for maximal oxygen consumption

Study	Exercise mode	Outcome	Group name	n	Pre-intervention Mean ± SD	Post-intervention Mean ± SD
Born et al. [72]	Running	$\dot{V}O_{2peak}$	HIIT	16	49.0 ± 4.5	51.5 ± 3.9
			LSD	12	52.4 ± 4.8	51.8 ± 5.3
Driller et al. [60]	Rowing	$\dot{V}O_{2peak}$	HIT	10	53.2 ± 6.4	57.0 ± 8.4
			CT	10	54.8 ± 7.3	54.2 ± 7.8
Esfarjani and Laursen [67]	Running	$\dot{V}O_{2max}$	G_1	6	51.3 ± 2.4	56.0 ± 1.4
			G_2	6	51.7 ± 3.4	54.9 ± 2.1
			G_{CON}	5	51.8 ± 2.8	52.9 ± 1.5
Ferley et al. [69]	Running	$\dot{V}O_{2max}$	G_{Flat}	12	59.4 ± 8.9	59.6 ± 7.6
			G_{Hill}	12	63.3 ± 8.0	62.7 ± 7.0
			G_{Con}	8	59.9 ± 8.6	58.3 ± 7.9
Hebisz et al. [61]	Cycling	$\dot{V}O_{2max}$	Endurance	13	57.9 ± 6.8	66.6 ± 5.3
			I	13	61.3 ± 7.5	62.4 ± 6.9
Hebisz et al. [70]	Cycling	$\dot{V}O_{2max}$	E1	10	57.4 ± 3.4	63.5 ± 6.1
			E2	7	58.6 ± 5.9	63.6 ± 3.9
			C	7	55.4 ± 7.2	58.9 ± 9.8
Hottenrott et al. [62]	Running	$\dot{V}O_{2peak}$	AW	14	36.8 ± 4.5	43.6 ± 6.5
			WE	16	38.8 ± 5.0	41.5 ± 6.0
Kirchenberger et al. [64]	Rowing	$\dot{V}O_{2peak}$	IG	10	58.4 ± 3.9	62.1 ± 3.9
			CG	7	58.4 ± 5.7	58.3 ± 6.4
Laursen et al. [73]	Cycling	$\dot{V}O_{2peak}$	HIT	7	68.7 ± 3.6	70.3 ± 5.7
			CON	7	66.3 ± 3.7	65.8 ± 4.3
Michalik et al. [65]	Running	$\dot{V}O_{2max}$	Experiment	6	52.7 ± 5.3	56.4 ± 3.3
			C	6	46.6 ± 4.7	50.9 ± 4.6
Seiler et al. [71]	Cycling	$\dot{V}O_{2peak}$	4 × 4 min	9	50.4 ± 5.8	53.2 ± 7.6
			4 × 8 min	9	52.8 ± 4.8	58.3 ± 5.8
			4 × 16 min	9	51.1 ± 5.8	54.4 ± 5.2
			Low	8	53.3 ± 6.8	55.9 ± 5.2
Silva et al. [68]	Running	$\dot{V}O_{2max}$	HIT	8	54.5 ± 8.1	57.1 ± 6.4
			CON	8	56.6 ± 7.3	56.9 ± 7.6

Results at pre and post are expressed in $mL \cdot kg^{-1} \cdot min^{-1}$

AW after work, *C, CG, CON,* G_{CON} control group, *CT* traditional training, *E1, E2* concurrent high-intensity interval training, sprint interval training, and continuous training, G_1 group 1, G_2 group 2, G_{Flat} level-grade interval training, G_{Hill} uphill interval training, *HIIT, HIT* high-intensity interval training, *I* interval and endurance training, *IG* intervention group, *Low* low-intensity, *LSD* long-slow distance, *n* sample size, *SD* standard deviation, $\dot{V}O_{2max}$ maximal oxygen consumption, $\dot{V}O_{2peak}$ maximal oxygen consumption, *WE* weekend

that a minimum of 81 participants per group (total of 162) would be required to obtain sufficient power to determine a significant effect (SMD 0.44) for $\dot{V}O_{2peak}$.

A surprising result was that there was no significant difference between groups at follow up testing for TT performance. However, the number of studies ($k = 5$) and the total number of participants ($n = 80$) included in the analysis were considered to be very low compared with the analysis for $\dot{V}O_{2peak}$. Furthermore, there was no statistical heterogeneity in the results for $\dot{V}O_{2peak}$, W_{peak}, or TT performance ($I^2 = 0.0\%$ for all), which is another surprising result. This implies that despite the superficially heterogeneous sample groups

Table 3 Results of individual studies for W_{peak}

Study	Exercise mode	Group name	*n*	Pre-intervention Mean ± SD	Post-intervention Mean ± SD
Driller et al. [60]	Rowing	HIT	10	352.2 ± 106.1	373.9 ± 102.5
		CT	10	358.2 ± 90.5	371.5 ± 98.9
Esfarjani and Laursen [67]	Running	G_1	6	18.4 ± 0.8	20.2 ± 0.7
		G_2	6	18.1 ± 0.6	20.2 ± 0.7
		G_{CON}	5	17.9 ± 0.5	18.3 ± 0.4
Ferley et al. [69]	Running	G_{Flat}	12	n/a	n/a
		G_{Hill}	12	n/a	n/a
		G_{Con}	8	n/a	n/a
Hebisz et al. [70]	Cycling	E1	10	421.2 ± 48.4	456.0 ± 45.2
		E2	7	392.8 ± 29.8	414.9 ± 40.2
		C	7	387.6 ± 49.6	396.6 ± 51.0
Hottenrott et al. [62]	Running	AW	14	11.9 ± 1.6	14.6 ± 2.2
		WE	16	13.0 ± 1.8	14.7 ± 2.0
Laursen et al. [73]	Cycling	HIT	7	469.0 ± 38.0	489.0 ± 45.0
		CON	7	490.0 ± 47.0	488.0 ± 45.0
Michalik et al. [65]	Running	Experiment	6	21.2 ± 1.2	21.7 ± 1.3
		C	6	18.8 ± 2.3	19.3 ± 2.3
Seiler et al. [71]	Cycling	4 × 4 min	9	356.4 ± 70.7	380.7 ± 75.7
		4 × 8 min	9	403.5 ± 39.4	446.9 ± 30.0
		4 × 16 min	9	380.4 ± 55.0	395.2 ± 53.5
		Low	8	370.4 ± 47.7	382.6 ± 52.3
Silva et al. [68]	Running	HIT	8	20.0 ± 2.2	21.2 ± 2.2
		CON	8	22.4 ± 1.3	22.1 ± 2.0
Stepto et al. [66]	Cycling	12 × 30 s	4	371.8 ± 28.6	373.3 ± 29.9
		12 × 60 s	3	349.7 ± 95.2	354.7 ± 91.6
		12 × 2 min	4	403.3 ± 21.0	411.0 ± 25.6
		8 × 4 min	4	389.8 ± 24.3	407.5 ± 26.0
		4 × 8 min	4	385.3 ± 18.2	390.3 ± 18.3

AW after work, *C, CG, CON,* G_{CON} control group, *CT* traditional training, *E1, E2* concurrent high-intensity interval training, sprint interval training, and continuous training, G_1 group 1, G_2 group 2, G_{Flat} level-grade interval training, G_{Hill} uphill interval training, *HIT* high-intensity interval training, *Low* low-intensity, *n* sample size, *n/a* not available, *SD* standard deviation, *WE* weekend, W_{peak} peak power output from an incremental exercise test for cycling in watts or running velocity in km·h^{-1}

in terms of participant characteristics, exercise mode, and training programming variables, these sources of variability did not influence the outcomes.

It is important to consider the influence of a learning effect when interpreting the results for TT performance. The coefficient of variation (CV) between two TT performance tests has been shown to be as high as 3.0 ± 2.9% in endurance-trained athletes when they have not completed a familiarization test [74]. Following a familiarization trial the CV is reduced to 0.9 ± 0.7% [74]. Since none of the studies in the analysis for TT performance included a familiarization trial, there may be greater variability in the result, limiting the ability to detect a significant difference between intervention groups. It is unlikely that a learning effect would influence the results for $\dot{V}O_{2max}$ since it is defined by pre-specified physiological criteria, such as a plateau in oxygen consumption despite an increase in work rate. $\dot{V}O_{2peak}$ may be limited by an individual's willingness and/or ability to complete the test. However, given that combining $\dot{V}O_{2max}$ and $\dot{V}O_{2peak}$ did not influence the results of the present study ($p = 0.39$), we conclude that it is unlikely that there was a generalizable learning effect for $\dot{V}O_{2peak}$ in the current study.

This systematic review can only speak to the efficacy of including high-intensity exercise in a single meso-cycle which may not provide sufficient stimulus for adaptations in TT performance. Previous findings suggest that incorporating high-intensity training throughout multiple meso-cycles as part of a periodized training program can lead to greater improvements in TT performance compared with non-periodized programs in trained individuals [75]. Interestingly, the same study did not observe

Table 4 Results of individual studies for time-trial performance

Study	Exercise mode	Distance (km)	Group name	n	Pre-intervention Mean ± SD	Post-intervention Mean ± SD
Driller et al. [60]	Rowing	2.0	HIT	10	437.0 ± 40.0	429.0 ± 40.0
			CT	10	434.0 ± 36.0	432.0 ± 38.0
Esfarjani and Laursen [67]	Running	3.0	G_1	6	679.6 ± 38.1	629.3 ± 28.8
			G_2	6	680.4 ± 31.0	657.4 ± 29.6
			G_{CON}	5	681.2 ± 34.4	680.4 ± 33.4
Ingham et al. [63]	Rowing	2.0	Mix	9	405.7 ± 8.1	400.0 ± 7.8
			Low	9	401.6 ± 10.2	393.7 ± 9.8
Kirchenberger et al. [64]	Rowing	2.0	IG	10	428.5 ± 33.4	425.6 ± 32.8
			CG	7	418.7 ± 31.3	413.5 ± 29.9
Silva et al. [68]	Running	5.0	HIT	8	1196.0 ± 173.0	1168.0 ± 135.0
			CON	8	1149.0 ± 153.0	1165.0 ± 164.0
Stepto et al. [66]	Cycling	40.0	12 × 30 s	4	3434.9 ± 209.7	3354.6 ± 165.0
			12 × 60 s	3	3618.4 ± 301.7	3608.2 ± 283.0
			12 × 2 min	4	3181.7 ± 39.3	3138.5 ± 106.0
			8 × 4 min	4	3356.4 ± 156.5	3258.8 ± 123.9
			4 × 8 min	4	3223.5 ± 166.7	3245.7 ± 169.8

Time-trial results at pre- and post-intervention are expressed in seconds

CG, CON, G_{CON} control group, *CT* traditional training, G_1 group 1, *G_2* group 2, *HIT* high-intensity interval training, *IG* intervention group, *Low* low-intensity, *Mix* mixed intensity, *n* sample size, *SD* standard deviation

additional gains in $\dot{V}O_{2peak}$ [75]. These combined findings could indicate that a structured periodized approach is not required for adaptive changes in $\dot{V}O_{2peak}$ but may be required to produce significant improvements in TT performance. This is consistent with a recent large scale analysis [76], case studies [77], and decades of observations in high performance environments that suggest different time scales for peak adaptation of $\dot{V}O_{2peak}$, fractional utilization, and gross efficiency with long-term endurance training.

4.2 Limitations of the Evidence Included in the Review

Several studies did not include a randomization process to allocate participants to their respective intervention groups [67–73]. The lack of randomization across studies could appear to have increased the possibility that participants' characteristics (e.g., age, sex, fitness, etc.) influenced the results in a biased manner [78]. In addition, only 2 of the 14 included studies stratified participants by baseline characteristics prior to group allocation [67, 68]. Matching participants by characteristics prior to randomization can be beneficial in sport science studies that have small sample sizes to prevent non-homogeneity among training groups [79]. However, an analysis for each outcome showed that there was no significant effect of the allocation process on the results or the pooled analyses.

If the effect for $\dot{V}O_{2peak}$ is the expected SMD between interventions, then the studies included in the analyses were underpowered (mean power of ~ 17%) in their ability to detect a significant difference between interventions. This suggests that the sample size of the included studies was too small, which is a common limitation in sport science research [80]. The power analysis performed in the current review showed that a minimum of 81 participants per group would be required to detect a truly significant difference between interventions of the magnitude suggested by the available studies. Larger sample sizes are required when comparing interventions in already trained subjects because the expected change in outcomes such as $\dot{V}O_{2peak}$ and TT performance following interval training in trained individuals is less than half the magnitude of that which occurs in untrained individuals [6, 16].

4.3 Limitations of the Review Processes

To account for differences in the measurement techniques used to program intensity among the studies (e.g., heart rate (HR), power/speed, $\dot{V}O_{2peak}$), the findings from Iannetta et al. [46] were used as a guide to classify interventions by intensity domain. It is important to note that the physiological responses to exercise are somewhat influenced by the mode of exercise (e.g., cycling, running) [81]. Since the study by Iannetta et al. [46] was conducted in cyclists, it is possible that intensity domains that were determined for the interventions in rowers and runners were imprecise. However, there was no evidence of statistical heterogeneity in the results for all three outcomes. Therefore, it is unlikely that the classification method influenced the results of the analysis.

When attempting to determine a causal relationship between an exercise intervention such as training above MMSS and changes in endurance performance, it is necessary to limit changes in the training program for the comparator group. In this systematic review, careful analysis of the training descriptions for the different groups revealed that 8 of 14 comparator (< MMSS) groups [60–62, 65–67, 70, 72] changed their training in some way during the intervention period. However, these changes were typically modest. Subgroup analyses showed that the changes had no significant impact on the results for $\dot{V}O_{2peak}$ ($p = 0.52$), W_{peak} ($p = 0.88$), or TT performance ($p = 0.94$) when compared with the groups that maintained previous training programs. Consequently, the difference between groups was likely a result of intensity-specific adaptations as opposed to 'training differently'.

4.4 Implications of the Results for Future Research

An important finding from the current review was that all the studies included in the pooled analysis were significantly under powered. As previously discussed, to obtain a sample of sufficient power (estimated 81 participants per group), sport science researchers may be required to fundamentally change the way training intervention studies are conducted. As daunting as this sounds, we should consider new approaches to increase participation such as conducting crowd sourced training intervention studies. This is increasingly feasible with the widespread adoption of training monitoring platforms used by individual athletes, and the emergence of virtual training and racing platforms available online that athletes can use from their homes, at all times of the year, where multiple data streams are logged for most participants [82]. We envision that hybrid intervention studies with both traditional laboratory and crowd-sourced components might provide an optimal mix of experimental control and statistical power.

A key to performing well designed training intervention studies is tight control of the training variables and potential confounders. Likewise, a key to summarizing and comparing intervention studies using meta-analytic methods is rigorous definition of the study inclusion criteria to ensure interpretations can be drawn from an appropriate pooling of similar literature, without excessive heterogeneity. This decision must be balanced against the imperative to translate these same findings back to real-world sport settings, where heterogeneity is the rule, not the exception. For example, to be included in this systematic review, studies had to include a training group that performed significant HIIT training as part of their weekly training as well as a training group that performed *only* training below the MMSS. In practice, observation of endurance athletes at all levels strongly suggests that both high-intensity training above MMSS and large volumes of low-intensity training are routinely combined in a coordinated and sustainable manner over time to drive maximal adaptation of responsive oxygen utilization cascade components [83–86].

4.5 Implications of the Results for Practice

The findings from the current review showed that training at intensities above MMSS led to greater improvement in $\dot{V}O_{2peak}$ in trained individuals. Athletes and coaches have known and followed this practice for many decades. Sport scientists, coaches, and athletes should incorporate high-intensity training bouts (i.e., intensity > MMSS) during training periods of at least 2 weeks in duration to obtain a significant improvement in $\dot{V}O_{2peak}$. The magnitude of difference in $\dot{V}O_{2peak}$ improvement for including training above MMSS versus training only below MMSS was modest, but may be meaningful in competition where results are determined by fractions of percent differences between athletes [87]. Our results suggested that there was no difference to improvement in TT performance or in W_{peak} with the addition of high-intensity training above MMSS.

The absence of statistical or clinical heterogeneity in the current findings, despite the apparent variety in the studies contributing to this meta-analysis is somewhat counter-intuitive. One possibility is that for athletes who are already trained in a specific sport, the addition of training intensity above MMSS matters more for further physiological adaptations than any other modifying variable. While training intensity appears to be important for $\dot{V}O_{2peak}$, and $\dot{V}O_{2peak}$ is important for performance outcomes, it is only one component of the integrated performance model along with measures of submaximal exercise efficiency [11]. In trained athletes, these submaximal components may not be as responsive to increasing the intensity of training within a single mesocycle. Previous findings have suggested that further increasing interval training intensity within the severe domain is less important for TT performance than increasing the duration of high-intensity training time [6]. While MMSS as the demarcation of heavy and severe intensity domains may be an important training threshold for physiological adaptations, it may not be as important for performance outcomes. This speculation should be interpreted with caution given the previously mentioned limitations in sample size and training cycle length available in the current literature.

A fundamental challenge for sport scientists attempting to offer evidence-based support for individual athlete training is translating from the traditionally reductionist, single mesocycle, and isolated training intervention literature to the integrated demands of a real-world training environment, with a process that stretches over multiple training phases and even quadrennial Olympic cycles. The average duration of the training intervention (for > MMSS groups) was 6 weeks, or 12–18 specific training sessions. When compared with subjects performing only low-intensity training, the net effect of these 12–18 training sessions (average 2.5 $mL \cdot kg^{-1} \cdot min^{-1}$ [~ 3–6%] increase in $\dot{V}O_{2peak}$) was substantial, both statistically and practically. However, it is less clear what the overall effect of an isolated intervention will be when considered *in media res* in an athlete's ongoing training. Indeed, the same intervention performed at different times in a season by the same athlete may yield markedly different adaptive results [88], suggesting the importance of considering the cumulative effects of training programming.

5 Conclusion

A single endurance training meso-cycle that includes regular training above MMSS can improve $\dot{V}O_{2peak}$ in already trained endurance athletes more than training below MMSS only. Despite the superficially heterogeneous sample groups in terms of participant characteristics, exercise mode, and training programming variables, these sources of variability did not influence changes in endurance performance. No significant difference related to intensity of training was observed for W_{peak} or TT performance after a single mesocycle. Although it may be well known that small sample size studies are prevalent in the field of sport science, the current review provided an indication of how substantial the issue is in peer-reviewed studies. To overcome these limitations, the utilization of technologies which allow for crowd sourced, digital data collection may be an important augmentation of the tools of sport science research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40279-023-01924-y.

Declarations

Registration and Protocol Not applicable.

Funding No sources of funding were used to assist in the preparation of this article.

Conflict of Interest Michael Rosenblat, Jem Arnold, Hannah Nelson, Jennifer Watt, and Stephen Seiler declare that they have no conflicts of interest relevant to the content of this review.

Availability of Data, Code and Other Material All data generated or analysed during this study are included in this published article (and its supplementary information files).

Author Contributions Michael Rosenblat conceived and designed the study, conducted article screening, data extraction, risk of bias analysis, statistical analysis, and manuscript preparation. Jem Arnold participated in data extraction and manuscript preparation. Hannah Nelson participated in article screening and risk-of-bias analysis. Jennifer Watt participated in the statistical analysis and manuscript preparation. Stephen Seiler participated in study design and manuscript preparation. All authors read and approved the final version of the manuscript.

References

1. Esteve-Lanao J, Juan AF, Earnest CP, Foster C, Lucía A. How do endurance runners actually train? Relationship with competition performance. Med Sci Sports Exerc. 2005;37(3):496–504.
2. Neal CM, Hunter AM, Galloway SD. A 6-month analysis of training-intensity distribution and physiological adaptation in Ironman triathletes. J Sports Sci. 2011;29(14):1515–23.
3. Tjelta LI, Enoksen E. Training characteristics of male junior cross country and track runners on European top level. Int J Sports Sci Coach. 2010;5(2):193–203.
4. Jones AM, Burnley M, Black MI, Poole DC, Vanhatalo A. The maximal metabolic steady state: redefining the "gold standard." Physiol Rep. 2019;7(10): e14098.
5. Poole DC, Jones AM. Oxygen uptake kinetics. Compr Physiol. 2012;2(2):933–96.
6. Rosenblat MA, Lin E, da Costa BR, Thomas SG. Programming interval training to optimize time-trial performance: a systematic review and meta-analysis. Sports Med. 2021;51(8):1687–714.
7. Hausswirth C, Louis J, Aubry A, Bonnet G, Duffield RO, Le Meur Y. Evidence of disturbed sleep and increased Illness in overreached endurance athletes. Med Sci Sports Exerc. 2014;46(5):1036–45.
8. Meeusen R, Duclos M, Foster C, Fry A, Gleeson M, Nieman D, et al. Prevention, diagnosis, and treatment of the overtraining syndrome: joint consensus statement of the European College of Sport Science and the American College of Sports Medicine. Med Sci Sports Exerc. 2013;45(1):186–205.
9. Roete AJ, Elferink-Gemser MT, Otter RTA, Stoter IK, Lamberts RP. A systematic review on markers of functional overreaching in endurance athletes. Int J Sports Physiol Perform. 2021;16(8):1065–73.
10. Rosenblat MA, Perrotta AS, Vicenzino B. Polarized vs. threshold training intensity distribution on endurance sport performance: a systematic review and meta-analysis of randomized controlled trials. J Strength Cond Res. 2019;33(12):3491–500.
11. Joyner MJ, Coyle EF. Endurance exercise performance: the physiology of champions. J Physiol. 2008;586(1):35–44.
12. Bassett DR, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc. 2000;32(1):70–84.

13. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. J Appl Physiol. 1955;8(1):73–80.
14. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. Med Sci Sports Exerc. 1995;27(9):1292–301.
15. Billat LV, Koralsztein JP. Significance of the velocity at VO2max and time to exhaustion at this velocity. Sports Med. 1996;22(2):90–108.
16. Rosenblat MA, Granata C, Thomas SG. Effect of interval training on the factors influencing maximal oxygen consumption: a systematic review and meta-analysis. Sports Med. 2022;52(6):1329–52.
17. Rosenblat MA, Perrotta AS, Thomas SG. Effect of high-intensity interval training versus sprint interval training on time-trial performance: a systematic review and meta-analysis. Sports Med. 2020;50(6):1145–61.
18. Granata C, Oliveira RS, Little JP, Renner K, Bishop DJ. Training intensity modulates changes in PGC-1alpha and p53 protein content and mitochondrial respiration, but not markers of mitochondrial content in human skeletal muscle. FASEB J. 2016;30(2):959–70.
19. MacInnis MJ, Zacharewicz E, Martin BJ, Haikalis ME, Skelly LE, Tarnopolsky MA, et al. Superior mitochondrial adaptations in human skeletal muscle after interval compared to continuous single-leg cycling matched for total work. J Physiol. 2017;595(9):2955–68.
20. Laursen PB, Shing CM, Peake JM, Coombes JS, Jenkins DG. Influence of high-intensity interval training on adaptations in well-trained cyclists. J Strength Cond Res. 2005;19(3):527–33.
21. Matsuo T, Saotome K, Seino S, Shimojo N, Matsushita A, Iemitsu M, et al. Effects of a low-volume aerobic-type interval exercise on VO2max and cardiac mass. Med Sci Sports Exerc. 2014;46(1):42–50.
22. Warburton DE, Haykowsky MJ, Quinney HA, Blackmore D, Teo KK, Taylor DA, et al. Blood volume expansion and cardiorespiratory function: effects of training modality. Med Sci Sports Exerc. 2004;36(6):991–1000.
23. Wright S, Esfandiari S, Elmayergi N, Sasson Z, Goodman JM. Left atrial functional changes following short-term exercise training. Eur J Appl Physiol. 2014;114(12):2667–75.
24. Esfandiari S, Sasson Z, Goodman JM. Short-term high-intensity interval and continuous moderate-intensity training improve maximal aerobic power and diastolic filling during exercise. Eur J Appl Physiol. 2014;114(2):331–43.
25. Huang YC, Tsai HH, Fu TC, Hsu CC, Wang JS. High-intensity interval training improves left ventricular contractile function. Med Sci Sports Exerc. 2019;51(7):1420–8.
26. Scribbans TD, Edgett BA, Vorobej K, Mitchell AS, Joanisse SD, Matusiak JB, et al. Fibre-specific responses to endurance and low volume high intensity interval training: striking similarities in acute and chronic adaptation. PLoS One. 2014;9(6): e98119.
27. Gist NH, Fedewa MV, Dishman RK, Cureton KJ. Sprint interval training effects on aerobic capacity: a systematic review and meta-analysis. Sports Med. 2014;44(2):269–79.
28. Milanović Z, Sporis G, Weston M. Effectiveness of high-intensity interval training (HIT) and continuous endurance training for VO2max improvements: a systematic review and meta-analysis of controlled trials. Sports Med. 2015;45(10):1469–81.
29. Wen D, Utesch T, Wu J, Robertson S, Liu J, Hu G, et al. Effects of different protocols of high intensity interval training for VO2max improvements in adults: a meta-analysis of randomised controlled trials. J Sci Med Sport. 2019;22(8):941–7.
30. Mattioni Maturana F, Martus P, Zipfel S, Nieß AM. Effectiveness of HIIE versus MICT in improving cardiometabolic risk factors in health and disease: a meta-analysis. Med Sci Sports Exerc. 2021;53(3):559–73.
31. PORBITAL. PlotDigitizer; 2023. https://plotdigitizer.com/app [cited 2023].
32. Morton RH. Why peak power is higher at the end of steeper ramps: an explanation based on the "critical power" concept. J Sports Sci. 2011;29(3):307–9.
33. Adami A, Sivieri A, Moia C, Perini R, Ferretti G. Effects of step duration in incremental ramp protocols on peak power and maximal oxygen consumption. Eur J Appl Physiol. 2013;113(10):2647–53.
34. Bergstrom HC, Housh TJ, Zuniga JM, Traylor DA, Camic CL, Lewis RW Jr, et al. The relationships among critical power determined from a 3-min all-out test, respiratory compensation point, gas exchange threshold, and ventilatory threshold. Res Q Exerc Sport. 2013;84(2):232–8.
35. Lucía A, Sánchez O, Carvajal A, Chicharro JL. Analysis of the aerobic-anaerobic transition in elite cyclists during incremental exercise with the use of electromyography. Br J Sports Med. 1999;33(3):178–85.
36. Greco CC, Carita RA, Dekerle J, Denadai BS. Effect of aerobic training status on both maximal lactate steady state and critical power. Appl Physiol Nutr Metab. 2012;37(4):736–43.
37. Galan-Rioja MA, Gonzalez-Mohino F, Poole DC, Gonzalez-Rave JM. Relative proximity of critical power and metabolic/ventilatory thresholds: systematic review and meta-analysis. Sports Med. 2020;50(10):1771–83.
38. Bergstrom HC, Housh TJ, Zuniga JM, Traylor DA, Lewis RW, Camic CL, et al. Responses during exhaustive exercise at critical power determined from the 3-min all-out test. J Sports Sci. 2013;31(5):537–45.
39. Brickley G, Doust J, Williams CA. Physiological responses during exercise to exhaustion at critical power. Eur J Appl Physiol. 2002;88(1–2):146–51.
40. McClave SA, LeBlanc M, Hawkins SA. Sustainability of critical power determined by a 3-minute all-out test in elite cyclists. J Strength Cond Res. 2011;25(11):3093–8.
41. Baron B, Noakes TD, Dekerle J, Moullan F, Robin S, Matran R, et al. Why does exercise terminate at the maximal lactate steady state intensity? Br J Sports Med. 2008;42(10):828–33.
42. Dittrich N, de Lucas RD, Beneke R, Guglielmo LG. Time to exhaustion at continuous and intermittent maximal lactate steady state during running exercise. Int J Sports Physiol Perform. 2014;9(5):772–6.
43. Fontana P, Boutellier U, Knöpfli-Lenzin C. Time to exhaustion at maximal lactate steady state is similar for cycling and running in moderately trained subjects. Eur J Appl Physiol. 2009;107(2):187–92.
44. Mendes TT, Fonseca TR, Ramos GP, Wilke CF, Cabido CE, De Barros CL, et al. Six weeks of aerobic training improves VO2max and MLSS but does not improve the time to fatigue at the MLSS. Eur J Appl Physiol. 2013;113(4):965–73.
45. Pethick J, Winter SL, Burnley M. Physiological evidence that the critical torque is a phase transition, not a threshold. Med Sci Sports Exerc. 2020;52(11):2390–401.
46. Iannetta D, Inglis EC, Mattu AT, Fontana FY, Pogliaghi S, Keir DA, et al. A critical evaluation of current methods for exercise prescription in women and men. Med Sci Sports Exerc. 2020;52(2):466–73.
47. Sterne JA, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366: l4898.
48. Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions; 2021. https://training.cochrane.org/handbook/current.
49. Hedges L, Olkin I. Statistical methods for meta-analysis. New York: Academic Press; 1981.

50. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1–48.
51. Gaddis GM, Gaddis ML. Introduction to biostatistics: Part 2, Descriptive statistics. Ann Emerg Med. 1990;19(3):309–15.
52. R Core Team. R: a language and environment for statistical computing. 4.2.2 ed. Vienna: R Foundation for Statistical Computing; 2022.
53. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.
54. Quintana DS. A guide for calculating study-level statistical power for meta-analyses. OSF Preprints. osf.io/js79t; 2022.
55. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum; 1988.
56. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.
57. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods. 2010;1(2):112–25.
58. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.
59. Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. Psychometrika. 1995;60(3):419–35.
60. Driller MW, Fell JW, Gregory JR, Shing CM, Williams AD. The effects of high-intensity interval training in well-trained rowers. Int J Sports Physiol Perform. 2009;4(1):110–21.
61. Hebisz P, Hebisz R, Zaton M, Ochmann B, Mielnik N. Concomitant application of sprint and high-intensity interval training on maximal oxygen uptake and work output in well-trained cyclists. Eur J Appl Physiol. 2016;116(8):1495–502.
62. Hottenrott K, Ludyga S, Schulze S. Effects of high intensity training and continuous endurance training on aerobic capacity and body composition in recreationally active runners. J Sports Sci Med. 2012;11(3):483–8.
63. Ingham SA, Carter H, Whyte GP, Doust JH. Physiological and performance effects of low- versus mixed-intensity rowing training. Med Sci Sports Exerc. 2008;40(3):579–84.
64. Kirchenberger T, Ketelhut S, Ketelhut RG. Effects of moderate- versus mixed-intensity training on VO2peak in young well-trained rowers. Sports. 2021;9(7):92.
65. Michalik K, Glinka S, Danek N, Zatoń M. Interval training with active recovery and the physical capacity of recreational male runners. Pol J Sport Tour. 2018;25(4):15–20.
66. Stepto NK, Hawley JA, Dennis SC, Hopkins WG. Effects of different interval-training programs on cycling time-trial performance. Med Sci Sports Exerc. 1999;31(5):736–41.
67. Esfarjani F, Laursen PB. Manipulating high-intensity interval training: effects on VO2max, the lactate threshold and 3000 m running performance in moderately trained males. J Sci Med Sport. 2007;10(1):27–35.
68. Silva R, Damasceno M, Cruz R, Silva-Cavalcante MD, Lima-Silva AE, Bishop DJ, et al. Effects of a 4-week high-intensity interval training on pacing during 5-km running trial. Braz J Med Biol Res. 2017;50(12): e6335.
69. Ferley DD, Osborn RW, Vukovich MD. The effects of uphill vs. level-grade high-intensity interval training on VO2max, Vmax, V(LT), and Tmax in well-trained distance runners. J Strength Cond Res. 2013;27(6):1549–59.
70. Hebisz R, Hebisz P, Borkowski J, Zatoń M. Effects of concomitant high-intensity interval training and sprint interval training on exercise capacity and response to exercise- induced muscle damage in mountain bike cyclists with different training backgrounds. Isokinet Exerc Sci. 2019;27(1):21–9.
71. Seiler S, Jøranson K, Olesen BV, Hetlelid KJ. Adaptations to aerobic interval training: interactive effects of exercise intensity and total work duration. Scand J Med Sci Sports. 2013;23(1):74–83.
72. Born D-P, Zinner C, Sperlich B. The mucosal immune function is not compromised during a period of high-intensity interval training. Is it time to reconsider an old assumption? Front Physiol. 2017;8:485.
73. Laursen PB, Blanchard MA, Jenkins DG. Acute high-intensity interval training improves Tvent and peak power output in highly trained males. Can J Appl Physiol. 2002;27(4):336–48.
74. Laursen PB, Shing CM, Jenkins DG. Reproducibility of a laboratory-based 40-km cycle time-trial on a stationary wind-trainer in highly trained cyclists. Int J Sports Med. 2003;24(7):481–5.
75. Bradbury DG, Landers GJ, Benjanuvatra N, Goods PS. Comparison of linear and reverse linear periodized programs with equated volume and intensity for endurance running performance. J Strength Cond Res. 2020;34(5):1345–53.
76. Almquist NW, Hansen J, Ronnestad BR. Development of cycling performance-variables and durability in female and male national team cyclists: from junior to senior. Med Sci Sports Exerc. 2023. https://doi.org/10.1249/MSS.0000000000003232.
77. Jones AM. The physiology of the world record holder for the women's marathon. Int J Sports Sci Coach. 2006;1(2):101–16.
78. Roberts C, Torgerson D. Randomisation methods in controlled trials. BMJ. 1998;317(7168):1301.
79. Laursen PB, Shing CM, Peake JM, Coombes JS, Jenkins DG. Interval training program optimization in highly trained endurance cyclists. Med Sci Sports Exerc. 2002;34(11):1801–7.
80. Heneghan C, Perera R, Nunan D, Mahtani K, Gill P. Forty years of sports performance research and little insight gained. BMJ. 2012;345: e4797.
81. Millet GP, Bentley DJ. Physiological differences between cycling and running: lessons from triathletes. Sports Med. 2009;39(3):179–206.
82. McIlroy B, Passfield L, Holmberg HC, Sperlich B. Virtual training of endurance cycling—a summary of strengths, weaknesses, opportunities and threats. Front Sports Act Living. 2021;3: 631101.
83. Guellich A, Seiler S, Emrich E. Training methods and intensity distribution of young world-class rowers. Int J Sports Physiol Perform. 2009;4(4):448–60.
84. Solli GS, Tonnessen E, Sandbakk Ø. The training characteristics of the world's most successful female cross-country skier. Front Physiol. 2017;8:1069.
85. Tønnessen E, Sylta Ø, Haugen TA, Hem E, Svendsen IS, Seiler S. The road to gold: training and peaking characteristics in the year prior to a gold medal endurance performance. PLoS One. 2014;9(7): e101796.
86. Zapico AG, Calderón FJ, Benito PJ, González CB, Parisi A, Pigozzi F, et al. Evolution of physiological and haematological parameters with training load in elite male road cyclists: a longitudinal study. J Sports Med Phys Fit. 2007;47(2):191–6.
87. Billat VL, Demarle A, Slawinski J, Paiva M, Koralsztein JP. Physical and training characteristics of top-class marathon runners. Med Sci Sports Exerc. 2001;33(12):2089–97.
88. Del Giudice M, Bonafiglia JT, Islam H, Preobrazenski N, Amato A, Gurd BJ. Investigating the reproducibility of maximal oxygen uptake responses to high-intensity interval training. J Sci Med Sport. 2020;23(1):94–9.

Authors and Affiliations

Michael A. Rosenblat[1] · **Jem Arnold**[2] · **Hannah Nelson**[3] · **Jennifer Watt**[4,5,6] · **Stephen Seiler**[7]

✉ Michael A. Rosenblat
michael@evidencebasedcoaching.ca
http://www.evidencebasedcoaching.ca

1 Evidence-Based Coaching, Toronto, ON, Canada

2 School of Kinesiology, Faculty of Education, University of British Columbia, Vancouver, BC, Canada

3 Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

4 Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada

5 Division of Geriatric Medicine, Department of Medicine, University of Toronto, Toronto, ON, Canada

6 Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

7 Department of Sport Science and Physical Education, Faculty of Health and Sport Sciences, University of Agder, Kristiansand, Norway